CALIFORNIA STATE UNIVERSITY
FULLERTON

CPSC 332 - File Structures and Databases
Project
Spring 2023

Important Date:
Part 1: Project Part A – Due May 03, 2023
Part 2: Project Part B – Due May 03, 2023

- Project submission should be made via Canvas. Each student should work on the project in groups of 4, and submit a final report, a demo, and presentation for the project.
- Each student in the group will be asked to submit a review of their group members, where each member will assign the percentage of contribution towards the project for their group members.
- Logically, an equal proportion of contribution from each member is expected.
- Any student who receives a review lower than 10% of other team members from their group members may receive a lower grade than other members.

Note: Project submissions made through email will not be accepted. Only submissions made on Canvas will receive grades. Only the group leader should submit PPT, report, and SQL file.

- Congratulations! You have been awarded a consultant database administrator job with the United Nations.
- The United Nations chief has collected some sample data on various countries around the world.
- She is now asking you to investigate the data that has been submitted from various countries.
- In particular, she is requesting that you investigate the database structure, create queries, and find interesting patterns in the data.

Project Part A [25 pts]:

Create an ER diagram for the world database that is provided in Canvas. This will require you to first examine the data to get a feel for the attributes for each table. Once you have a feel for the data, create the following. Please note, the database may not have foreign keys designated, so think through what foreign keys may be necessary to ensure referential integrity. Submit one ER Diagram and one Physical Model (one per group.) Make sure the following are shown in the ER Diagram:

- You will use the World Database from the LinkedIn Assignments as the **3 base tables (script creation file on Canvas)**
- Add **3 or more** additional tables to the base file (10 points)
    - You will need to get additional data that is relevant to the base tables.
    - The additional data can be dummy data but should make sense.
- Primary keys (5 points)
- Foreign keys (5 points)
- Relationships and any integrity constraints that make sense e.g., not null, date should be < today's date etc. (5 points)

Project Part B [50 pts]:

- The United Nations chief has asked you to develop 20 queries. These queries should be innovative (not currently covered in any training manuals).
- In particular, from the 20 queries, the chief has asked you to deliver at least 5 queries that must be structured with joins (10 points) and at least 5 structured with subqueries (10 points), and the remaining can be your choice (5 points).
- Create a final report that covers the following below. At the end of the semester, you will present your findings along with your team to the class in about a 15 - minute presentation. (25 points)
- You will need to find relevant data to add to the additional tables. It should make sense with the real world e.g., a table titled: **customer** should contain **customer** information.

Your final report will be in a **Word document format** and will outline the process of taking the requirements and showing the design process of the database. The report should be in APA style, 12-point font, double-spaced, Times New Roman, at least 4 pages excluding the title page and reference page. The recommended outline is as follows:

- **Title (Project Team Name, Class, Date, Team Members)**
- **Introduction**
  - Look at the data included in the database
  - Lookup the data in the UN website or various sites and explain its importance
  - Why does this database matter? E.g., GNP and life expectancy matter?
- **Database Development Process**
  - ER Diagram (from part A)
  - Explain your though process for the ER diagram
  - Database schema with primary and foreign keys identified
  - Logical and Physical Relational Models
- **Database Investigation**
  - SQL Queries
  - Include 20 as mentioned above; write out each one using production SQL syntax
  - At least 5 queries should have joins from multiple tables
  - At least 5 queries should have subqueries from multiple tables
  - Explain the purpose of each query and the value of it
- **Summary**
  - Provide a summary of the project
  - Example: pitfalls, revelations, etc.

Summary of Points:

3 Additional Tables	10 points
Primary Keys	5 points
Foreign Keys	5 points
Relationships	5 points
Min. 5 queries with joins	10 points
Min 5 queries with subqueries	10 points
Remaining queries	5 points
Report & Presentation	25 points